



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Identifying heterogeneity in rates of morphological evolution

**Citation for published version:**

Lloyd, GT, Wang, SC & Brusatte, SL 2012, 'Identifying heterogeneity in rates of morphological evolution: Discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi)', *Evolution: International Journal of Organic Evolution*, vol. 66, no. 2, pp. 330-348. <https://doi.org/10.1111/j.1558-5646.2011.01460.x>

**Digital Object Identifier (DOI):**

[10.1111/j.1558-5646.2011.01460.x](https://doi.org/10.1111/j.1558-5646.2011.01460.x)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Evolution: International Journal of Organic Evolution

**Publisher Rights Statement:**

Published version is available copyright of Wiley-Blackwell (2012) available at [www.interscience.wiley.com](http://www.interscience.wiley.com)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



**Author Post-Print Version.** Final version available in *Evolution* published by Wiley-Blackwell (2012).

Cite As: Lloyd, GT, Wang, SC & Brusatte, SL 2012, 'Identifying heterogeneity in rates of morphological evolution: Discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi)' *Evolution*, vol 66, no. 2, pp. 330-348.

DOI: 10.1111/i.1558-5646.2011.01460.x

## **Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi)**

Graeme T. Lloyd<sup>1,2</sup>, Steve C. Wang<sup>3,4</sup> and Stephen L. Brusatte<sup>5,6,7</sup>

<sup>1</sup>Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK

<sup>2</sup>E-mail: [graemetlloyd@gmail.com](mailto:graemetlloyd@gmail.com)

<sup>3</sup>Department of Mathematics and Statistics, Swarthmore College, Swarthmore, PA 19081, USA

<sup>4</sup>E-mail: [scwang@swarthmore.edu](mailto:scwang@swarthmore.edu)

<sup>5</sup>Division of Paleontology, American Museum of Natural History, Central Park West at 79<sup>th</sup> Street, New York, NY 10024, USA

<sup>6</sup>Department of Earth and Environmental Sciences, Columbia University, New York, NY, USA

<sup>7</sup>E-mail: [Stephen.brusatte@ed.ac.uk](mailto:Stephen.brusatte@ed.ac.uk) now of the University of Edinburgh, School of Geosciences, UK.

Abstract: Quantifying rates of morphological evolution is important in many macroevolutionary studies, and critical when assessing possible adaptive radiations and episodes of punctuated equilibrium in the fossil record. However, studies of morphological rates of change have lagged behind those on taxonomic diversification, and most authors have focused on continuous characters and quantifying patterns of morphological rates over time. Here, we provide a phylogenetic approach, using discrete characters and three statistical tests to determine points on a cladogram (branches or entire clades) that are characterized by significantly high or low rates of change. These methods include a randomization approach that identifies branches with significantly high rates and likelihood ratio tests that pinpoint either branches or clades that have significantly higher or lower rates than the pooled rate of the remainder of the tree. As a test case for these methods we analyze a discrete character dataset of lungfish, which have long been regarded as “living fossils” due to an apparent slow down in rates since the Devonian. We find that morphological rates are highly heterogeneous across the phylogeny and recover a general pattern of decreasing rates along the phylogenetic backbone towards living taxa, from the Devonian until the present. Compared with previous work, we are able to report a more nuanced picture of lungfish evolution using these new methods.

KEY WORDS: adaptive radiation, evolutionary rates, lungfish, morphological evolution, paleontology, punctuated equilibrium

“Evolutionary rate” can refer to many different concepts, ranging from the rate of diversification (speciation minus extinction; e.g., Ricklefs 2007) to the pace of molecular sequence change (e.g., Mindall & Thacker 1996). One rate concept of particular interest is the rate of morphological change. In other words, how rapidly do features of the phenotype change in different groups of organisms or during different time periods? Such questions are central to the long-standing debate regarding gradual and punctuational modes of evolutionary change (e.g., Eldredge & Gould 1972). Similarly, understanding the tempo of phenotypic change is essential for answering many macroevolutionary questions. Do some groups, the so-called “living fossils” of textbook fame, truly change very little over great periods of time? Are some groups characterized by a heritable shift, either a decrease or increase, in morphological rate relative to close relatives? Are high rates of change concentrated early in a group’s history, perhaps in concert with an adaptive radiation (e.g., Gavrillets & Losos 2009)? And what organismal features or evolutionary processes may promote high or low rates of change?

Before interpreting the causes and effects of high or low rates of morphological change, it is necessary to rigorously quantify the pattern of rates within a group or time period of interest. Are rates actually heterogeneous—significantly different from a pattern of randomly uniform morphological change—among the organisms or times in question? If so, exactly which groups or time periods are characterized by a significantly high or low rate of change? Furthermore, can we identify specific points on a phylogenetic tree, either specific branches or entire clades, that are characterized by higher or lower than expected rates of change? These questions are analogous to those often posed in the literature on diversification over time (e.g., Chan & Moore 2002; Sims & McConway 2003; McConway & Sims 2004; Ree 2005; Ricklefs 2007; Alfaro et al. 2009; Moore & Donoghue 2009). Biologists are often interested in whether specific “key adaptations” may have caused clades

to become more diverse than closely-related groups, but as a first step must establish that these clades are actually more diverse than expected by random chance.

Most previous studies have quantified morphological rates using morphometric or other continuous character data, such as body size or skull shape (e.g., Garland 1992; Hutcheon & Garland 2004; Collar et al. 2005; O'Meara et al. 2006; Sidlauskas 2007; Harmon et al. 2008; Pinto et al. 2008; Adams et al. 2009; Cooper & Purvis 2009). Discrete characters, such as those commonly used in morphological phylogenetic analyses, are rarely marshalled as primary data in evolutionary rates studies, with only a few examples in the literature (e.g., Westoll 1949; Derstler 1982; Forey 1988; Ruta et al. 2006; Brusatte et al. 2008). This is a curious situation as large databases of discrete characters, associated with well-resolved phylogenetic trees, are available for many groups of organisms. Furthermore, discrete characters can easily be optimized onto a tree, and rates of change for individual branches can be calculated by dividing the number of characters changing along that branch by the time duration of the branch (Wagner 1997). This procedure has been explored by a handful of authors, but has mostly been used to compare morphological to molecular rates of change along branches (e.g., Omland 1997; Bromham et al. 2002), or to calculate temporal curves of evolutionary rates (e.g., Wagner 1997; Ruta et al. 2006; Brusatte et al. 2008). Such temporal curves are useful for exploring the relationship between rates and major geological (e.g., mass extinctions) or biological (e.g., colonization of new ecospace) events, but do not answer the question of whether specific groups of organisms (branches or clades) have heterogeneous rates that differ from random uniformity.

Here, we use discrete characters, examined in a phylogenetic context, to determine exact points on a cladogram (branches or entire clades) that are characterized by significantly high or low rates of change. In other words, we address whether specific branches linking two taxa have rates of change significantly different from random variation, or whether entire

clades are characterized by higher or lower rates of change than the remainder of the tree. The latter question is particularly interesting, as these clades may be described as having a “rate shift,” potentially indicative of a heritable change in rate that is retained in descendant taxa (i.e, members of that clade). To determine significant rate deviations we use a likelihood-based method. Such methods have been used previously in several studies of continuous morphological change (e.g., Pagel 1998; Oakley 2003; Collar et al. 2005; O’Meara et al. 2006; Thomas et al. 2006; Revell et al. 2008), and are ideal in the present case because of their flexible modelling capabilities and tractable statistical properties.

As a case study for these methods we examine lungfish (Sarcopterygii; Dipnoi), a well-known group with a long history of analytical study with respect to their rate of morphological evolution (Westoll 1949; Simpson 1953; Martin 1987; Schultze 2004). Lungfish have also played a prominent historical role in the development of evolutionary thought, as a perceived pattern of high rates of change early in their history, and a subsequent decrease toward the present, led Darwin (1859) to refer to lungfish as “living fossils.” In fact, lungfish are still among the classic textbook examples of “living fossils” (literally—see for example: Simpson 1953, Raup & Stanley 1978, Stanley 1979, Ridley 1996, Levinton 2001), and often receive mention in popular accounts (Jones 1999, Dawkins 2004). Such groups have apparently changed very little over great periods of time, though this notion has rarely been tested. It is therefore important to determine whether lungfish do show decreasing rates of character change in certain regions of their phylogeny, and whether specific lungfish clades are characterized by a significantly lower rate of change than others.

## Materials and Methods

### *Dataset*

A number of cladistic data sets exist for lungfish (Marshall 1986; Schultze & Marshall 1993; Schultze & Chorn 1997; Krupina & Reisz 2000; Schultze 2001, 2004; Ahlberg & Smith 2006; Apesteguia et al. 2007; Cavin et al. 2007; Friedman 2007; Qiao & Zhu 2009). However, many concentrate solely on those taxa that lived during the earliest phase of lungfish history in the Devonian, a time thought to be characterized by high rates of change (Krupina & Reisz 2000; Schultze 2001; Ahlberg & Smith 2006; Friedman 2007; Qiao & Zhu 2009), or more recent post-Devonian taxa that have apparently changed at a more gradual rate (Schultze 2004; Apesteguia et al. 2007; Cavin et al. 2007). For our study, we require a global phylogenetic dataset that includes taxa from throughout lungfish history, to more properly test whether rates of character change have differed at different times and in different clades. Thus, we utilize the dataset of Schultze and Marshall (1993), which includes 60 ingroup taxa of Early Devonian–Recent age scored for 90 morphological characters. This dataset has been modified for more recent phylogenetic analyses focusing on Devonian (Schultze 2001) and Mesozoic (Schultze 2004) taxa.

We have combined the original dataset with the two recent modifications, along with the smaller matrix of Schultze & Chorn (1997), into a “supermatrix” (e.g., Gatesy et al. 2002, 2004; Hill 2005). Two taxonomic changes and one character modification were made prior to tree searches. *Iranorhynchus* was removed, as this genus has dubious dipnoan affinities (Marshall 1986) and can only be coded for three characters. We inserted *Amadeodipterus* (Young & Schultze 2005), which was recently described in accordance with the same coding scheme. For clarity’s sake, multistate character 36 (presence of cheek bones 11 and 10) was

split into two binary presence-absence characters (although this did not affect the phylogenetic result). This final and enlarged dataset includes two outgroup (*Psarolepis romeri* and *Diabolepis speratus*) and 84 ingroup species, comprising at least one representative of nearly all known dipnoan genera, scored for 91 characters (see Electronic Supplementary Material).

### *Obtaining a resolved cladogram*

The rates analysis we present requires a fully resolved cladogram, onto which characters can be optimized. As our supermatrix is a novel dataset that has not previously been analyzed, we performed a phylogenetic analysis in PAUP\* (Swofford 2003). Due to the size of the dataset tree searches were performed using the parsimony ratchet (Nixon 1999) implemented by creating a PAUP\* command file using perlRat.pl (Bininda-Emonds 2004). Twenty batches of 1,000 replicates were run, which returned 20,000 trees. These were then filtered to find the shortest tree, and duplicates were removed. These trees were then used as a starting point for a heuristic search to recover other most parsimonious trees, which returned 784,000 trees before the memory limit was reached. Consensus trees show that the lack of resolution is strongly concentrated amongst the Devonian taxa. As it was not feasible to analyze all trees, and the sheer size of both the dataset and resulting trees prevented a tractable reduced consensus method (Wilkinson 1994), a more pragmatic approach was sought. Here, a majority rule approach was used that involved iteratively producing majority-rule consensus trees and pruning out trees that were incongruent with it. At conclusion this process produced just two trees, although as they produced nearly identical results only the first is reported. This method was preferred over other methods such as, for example, using stratigraphic fit to choose between trees, which would bias the results towards higher rates (as



this would necessitate the sum of branch durations – the divisor in a rate equation – to be minimized). Nevertheless the variability between phylogenetic hypotheses should not be ignored, so we chose 10 equally spaced additional trees and repeated the tests outlined below in each case, although only using 100, rather than 1,000 iterations (see below). As the results for these trees do not affect the conclusions drawn below we confine these results to the supporting information.

### *Variation in character optimization*

A single cladogram may still vary in the evolutionary rates it implies depending on the way characters are optimized on the tree (i.e., the mapping of what changes occur on which branches), which directly affects branch lengths (number of character changes per branch). Typical optimizations in a parsimony framework, which is the most common method of phylogeny reconstruction for morphological datasets, are accelerated (ACCTRAN) or delayed (DELTRAN) transformation (Swofford & Maddison 1987). Accelerated optimization favors losses over convergence, and places character changes toward the root of a tree. Delayed optimization, on the other hand, favors convergence over losses and places changes toward the tips of the tree. There are other optimization procedures, involving maximum likelihood and Bayesian techniques, but these are less frequently used for discrete morphological datasets. In all cases we count the number of changes in the standard cladistic way, so for example an unordered character change from 0 to 2 is 1 change, but the same shift in an ordered character is 2 changes.

ACCTRAN and DELTRAN result in a single discrete state optimized for each character at each node. Ideally, it would be beneficial to investigate the entire range of possible character changes across the tree, but this would require varying optimizations

character-by-character, a procedure that would demand an inordinately large number of runs (and hence greatly increase the length of computing time). Here we favor a more pragmatic approach of examining two disparate character optimizations, one where all ambiguous changes are plotted according to ACCTRAN and the other where all changes were plotted according to DELTRAN. These two methods represent approximate end members of the full spectrum of parsimony optimization values, and this approach has a precedent in the literature (e.g., Brusatte et al. 2008). (Surprisingly, other previous work on lungfish rates did not explicitly address this issue.) Future studies, of course, could investigate other optimization procedures, including likelihood and Bayesian techniques.

#### *Problems with missing data*

Phylogenetic methods such as those employed here are sensitive to two potential biases associated with missing data: 1) incompletely known taxa, those for which only a proportion of characters can be coded; and 2) unsampled taxa, those which cannot be included in an analysis either because they are not available for study or were never preserved as fossils. These two biases are likely to affect the results in fundamentally different ways. The former case is explicitly accounted for in the methods outlined below, but the latter currently has no implemented solution, because it is impossible to predict the existence, and more importantly the character scores and temporal duration, of unsampled taxa. This problem is not unique to the current study—any phylogenetic, biogeographic, molecular clock, or disparity analysis is faced with the same issue—and is accepted as an unfortunate and unavoidable reality by biologists and paleontologists. This problem, however, should not be considered fatal. The methods presented here are general techniques that can be applied to any empirical dataset, and the results are contingent upon the reality of

the group being studied. As an analogy, general phylogenetic methods such as parsimony, likelihood, and Bayesian techniques are valid and useful, but may give incorrect answers if the character data is faulty or important taxa are not included in the analysis.

In the face of this reality, it is important to demonstrate two things. First, in reference to our chosen empirical test case, we provide a quantitative assessment of the quality of the lungfish fossil record through time. If there are systematic gaps in the fossil record—in this case, entire time intervals that are poorly sampled relative to other intervals—then differential sampling alone may produce a signal similar to elevated and/or depressed rates of change during intervals that have better/worse fossil records. If differential sampling does exist, then this is critical to keep in mind when interpreting the results of the rate analysis. Therefore, researchers utilizing our methods should always be careful to consider sampling. We provide a quantitative consideration of the lungfish fossil record below, and use this information to guide interpretation of our results. Second, we also use simulations (below) to explore the sensitivity of our methods to different levels of biased sampling. These results are useful in helping researchers decide whether an empirical group they are studying is known from a good enough fossil record for insightful application of our morphological rates methods.

We also note two potential protocols for mitigating the effects of biased sampling, although we employ neither here, because we demonstrate that the quality of the lungfish fossil record is generally good. The first would be to simply randomly delete taxa and repeat the tests to gauge any effects. A major drawback with this approach is its time consuming nature. Most of the computational tests used here take hours or days to run, and these would be supplemented with additional tree searching time (which may also take hours or days). Thus, a random deletion approach may only be appropriate for very small datasets at this time. A second approach would be to sample evenly through time to begin with (by taking a fixed number of taxa from each time bin), such that, although not eliminated, the problem is

not concentrated at particular time periods. However, this approach would likely be prone to type II errors since potentially important data would be discarded.

#### *Dating trees of fossil taxa: sensitivity test*

We date our trees following the method of Ruta et al. (2006; see Brusatte 2011 and ESM for a full description of this method). The time durations of each branch of the tree are calculated based on the absolute ages of the terminal taxa. However, this introduces a source of ambiguity, because terminal taxa in the fossil record are rarely dated with certainty. Each taxon date, which is essentially the date of its entombing rocks as determined by radioisotopic dating or correlation with dated units, is always associated with error bars. We agree with Pol & Norell (2006) that it is no longer acceptable to assume that dates are known with certainty, especially in the case of vertebrate fossils which are often only assignable to the stage or epoch level and thus imply an uncertainty of several million years. In order to counter this problem a randomization approach, similar to that of Pol & Norell (2006), was adopted, in which numerical dates were assigned to each taxon by drawing at random from a uniform distribution bounded by the error bars associated with each taxon age. Here, these error bars constitute the time bounded by the base and top of the stage or epoch, using dates from Gradstein *et al.* (2004), from which the first appearance of that taxon is known in the observed fossil record. In order to assess the sensitivity of our test to this uncertainty, this approach was repeated 1000 times, and in all cases the full spread of results is reported.

#### *Three evolutionary rates tests*

With a list of discrete characters, a fully resolved phylogenetic tree, and an age range

for each taxon, it is now possible to calculate evolutionary rates for each branch of the tree, and to test whether certain clades or time periods are characterized by statistically significant rate deviations from the remainder of the tree. Once again, in this approach, evolutionary rate is defined as the number of characters changing along a branch divided by the time duration of that branch. We outline three separate tests that, depending on the practical realities of the dataset at hand and the specific evolutionary questions being asked, are useful for determining rate heterogeneity across a tree. The first of these tests is a simple randomization that identifies specific branches that have unusually high rates of change. The second test is a likelihood ratio test that assesses whether specific branches have significantly high or low rates of change, and the third is a likelihood ratio test that assesses whether entire clades have higher or lower rates of change than the remainder of the tree. All tests were performed in R (R Development Core Team 2009) using the APE library (Paradis et al. 2004) and the code is available on request from the authors.

#### *Branch randomization test*

One frequent and straightforward question is whether or not we can identify branches whose rate of character change is significantly different (usually higher) than expected. In this case, “significantly different than expected” refers to a difference from a rate calculated from a tree in which branch lengths (the number of character changes) were assigned at random, assuming that the chances of a change occurring along a branch is equal at all times and all places across the tree. This approach draws upon a dated tree (using the method described above) as a starting point. The randomization proceeds by effectively laying all of these branches end-to-end and rescaling them so that their total summed duration is equal to 1. Random numbers between 0 and 1 are then drawn from a uniform distribution and

assigned to a specific branch based on this sequence. This process is repeated for the total number of character changes observed on the real tree; thus, the process mimics the case where the probability of a character change is equal at all times and in all places. Consequently, each branch may have anywhere between zero and the total number of changes assigned to it. Repeating this entire process 1,000 times gives a distribution of changes for each branch, and then if the actual number of changes observed is greater than that in 950 of the iterations then it is deemed significantly high. (A one-tailed test – looking for significantly *high* rates only – was chosen here as the overall rate of character change in our lungfish dataset was sufficiently low that the lowest possible value, zero, so often occurred that it was almost impossible to get a significantly low value.) However, the method can in principle be used to detect branches with significantly low rates in taxa with higher rates of morphological change.

One minor modification to this process was made following the observation of Wagner (1997) that terminal branches leading to incompletely known taxa (e.g., those known from only a skull and not the entire skeleton) necessarily carry a lower chance of recording a change, due to missing data in the terminals alone. Consequently some terminal branches were rescaled by dividing the number of sampled characters by the total number of characters so that the chance of a random change being assigned was truly equal across the tree.

### *Branch likelihood test*

An alternative to the randomization approach is to use likelihood-based approach, which we implement as follows. Let  $k$  denote the number of branches in the tree. For branch  $i$  ( $i = 1 \dots k$ ), let  $X_i$  denote the number of character changes occurring along that branch; let  $t_i$  denote the time duration of the branch in millions of years; let  $\lambda_i$  denote the intrinsic rate of

evolution of the branch; and let  $c_i$  denote the observed completeness of the branch (the proportion of the total number of characters that can be evaluated, based on the fossil material known). We model  $X_i$  as a Poisson process with rate parameter  $\lambda_i$ . Here  $\lambda_i$  can be thought of as the expected number of character changes per lineage million years if all characters were observable (i.e., if  $c_i = 1$ ). Then, the probability of observing  $x$  changes is given by the probability mass function

$$P(X_i = x | \lambda_i) = \frac{e^{-\lambda_i t_i c_i} (\lambda_i t_i c_i)^x}{x!}$$

The expected number of changes in branch  $i$  is thus  $\lambda_i t_i c_i$ . For example, suppose  $\lambda_i = 3$  changes per million years, the branch has a duration of 4 million years, and 50% of the total number of characters can be observed. We would then expect an average of  $(3)(4)(.5) = 6$  observed character changes in this branch.

The Poisson process model is appealing because of its simplicity, wide applicability, and prior usage in the macroevolution literature (e.g., Ree 2005). The model has several inherent assumptions, notably that character changes are reconstructed without error; characters along a branch are defined such that they evolve independently of each other; multiple character changes cannot occur at simultaneous instants in time (although multiple changes can occur along the same branch); the probability of a character change is uniform along each branch; and taxon sampling is uniform. We further assume, in calculating the joint likelihood, that each branch is independent of other branches. Although these assumptions cannot be tested directly, they should be reasonable if the characters under study are chosen appropriately.

We use a likelihood ratio test (LRT) to test hypotheses of interest. This test is based on the likelihood ratio (LR):

$$\text{LR} = \frac{\text{value of likelihood function under the null hypothesis}}{\text{maximum possible value of likelihood function under the null or alternative hypotheses}}$$

In our case, the likelihood function  $l(\lambda_1 \dots \lambda_k \mid x_1 \dots x_k)$  is given by

$$l(\lambda_1 \dots \lambda_k \mid x_1 \dots x_k) = \prod_{i=1}^k l(\lambda_i \mid x_i) = \prod_{i=1}^k P(X_i = x \mid \lambda_i) = \prod_{i=1}^k \frac{e^{-\lambda_i t_i c_i} (\lambda_i t_i c_i)^{x_i}}{x_i!}$$

For instance, suppose we wish to test the null hypothesis that all branches of the tree have a single common rate parameter, i.e.,  $\lambda_1 = \lambda_2 = \dots = \lambda_k$ , against the alternative hypothesis that not all rates are equal. Under the null hypothesis, all branches have equal intrinsic rates of evolution, and the numerator of the LR is the likelihood evaluated at  $\lambda_1 = \lambda_2 = \dots = \lambda_k = \hat{\lambda}$ , where  $\hat{\lambda} = \sum x_i / \sum t_i c_i$  is the maximum likelihood estimate obtained by pooling data from all branches. The denominator is given by the likelihood evaluated at  $\lambda_1 = \hat{\lambda}_1, \lambda_2 = \hat{\lambda}_2, \dots, \lambda_k = \hat{\lambda}_k$ , where  $\hat{\lambda}_i = x_i / t_i c_i$  is the maximum likelihood estimate for branch  $i$  alone. The test statistic for the LRT is  $-2\log(\text{LR})$ , which approximately follows a chi-square distribution with  $k - 1$  degrees of freedom under the null hypothesis, from which the p-value for the test can be calculated.

We also use an LRT framework to identify branches that significantly differ in their rate of evolution from the rest of the tree. This is equivalent to testing the null hypothesis  $\lambda_1$



$= \lambda_2 = \dots = \lambda_k$  against the alternative hypothesis  $\lambda_i \neq \lambda_1 = \lambda_2 = \dots = \lambda_{i-1} = \lambda_{i+1} \dots = \lambda_k$  for each branch  $i$ . It is straightforward to modify the LRT described above to test this hypothesis: the numerator remains the same, and the denominator is now the likelihood evaluated at  $\lambda_i = \hat{\lambda}_i = x_i / t_i c_i$  (the maximum likelihood estimate for branch  $i$  alone) and  $\lambda_1 = \lambda_2 = \dots = \lambda_{i-1} = \lambda_{i+1} \dots = \lambda_k = \hat{\lambda}_{(-i)} = \sum_{j \neq i} x_j / \sum_{j \neq i} t_j c_j$  (the maximum likelihood estimate obtained by pooling data from all branches except branch  $i$ ). Here the test statistic  $-2\log(\text{LR})$  follows a chi-square distribution with 1 degree of freedom under the null hypothesis. To account for multiple comparisons in testing multiple branches, we used the procedure of Benjamini and Hochberg (1995) to control the false discovery rate (FDR). This relatively recent method controls the expected proportion of falsely rejected hypotheses rather than the family-wise error rate (the latter being the overall probability of committing any false rejections). In our situation the FDR method is preferable to approaches such as the commonly used Bonferroni correction, as the latter tends to be overly conservative when the number of comparisons is large, as is the case here.

### *Clade likelihood test*

A second and related question to that posed above is whether there are specific clades on a tree that, as a whole, exhibit a significantly higher or lower rate than the remainder of the tree. If so, then in some cases these clades may be interpreted as evolving a heritable rate shift at or near their base, which is retained in descendants. In other words, this approach may test the possibility that rates of change can be inherited. Furthermore, this is an attempt to produce a morphological equivalent of the test performed by the software SymmeTREE (Chan & Moore 2005), which assesses the shape of a phylogeny to test for significant shifts

in the rate of diversification (the interplay of speciation and extinction of lineages; see below).

For this test, all non-terminal, non-root nodes were identified, as these define clades that can be compared with another portion of the phylogeny (i.e., the rest of the tree). A rate was then assigned to each portion (the clade and non-clade) by simply summing the number of character changes and dividing by the total amount of time spanned by the constituent branches. These two values represent a two-parameter model of evolutionary change that can then be compared with a single parameter model (one rate across the entire tree) using a likelihood ratio test. Such a test is analogous to the branch likelihood test described above, except that the denominator is now evaluated at the maximum likelihood estimate for the set of branches in the clade and the set outside the clade. As with the branch likelihood test, missing characters are accounted for using the  $c_i$  coefficients.

As the branch and clade likelihood tests depend on an asymptotic chi-square approximation, we ran simulations to verify that the tests worked correctly for sample sizes and tree structure equal to that of our dataset. (It was not necessary to run similar simulations for the randomization-based test, which is nonparametric and does not depend on an asymptotic approximation.)

#### *Character changes per lineage million years through time*

Aside from identifying specific lineages or clades with significantly high or low rates of morphological evolution, it is also interesting to determine if rates are heterogeneous over time (e.g., Forey 1988, Wagner 1997; Ruta et al. 2006, Brusatte et al. 2008). However, it is not straightforward to ascertain when a “rate” occurs, as a branch spans a length of time instead of pinpointing a discrete moment in time. In other words, if a branch spans 25 million

years, then how should the rate of change along that branch be binned? In previous studies, Ruta et al. (2006) simply counted a “rate” as occurring where the branch begins (i.e., at the earliest possible point, the time of the first sampled descendant) and Brusatte et al. (2008) followed the same procedure.

Here a slightly different approach is used (following Chaloner & Sheerin 1979), whereby instead of ascertaining when a rate (change/time) occurs, an attempt is made to measure when individual changes could have occurred in time. Thus, this approach focuses on individual discrete character changes instead of per-branch rates. This was done by drawing random numbers (equal in number to the changes along a branch) from a uniform distribution between the start of a branch and its end, resulting in a vector of ages for each character change. So, for example, if a branch spanned between 20 and 10 Ma and five changes were known to occur along it, then 5 numbers were drawn from a uniform distribution with a minimum of 10 million and a maximum of 20 million. The summed number of changes in a bin (here we use the c. 11-million year bins of Alroy et al. 2008) was then divided by the summed duration of the branches within that bin to get a per lineage million years rate. Accumulation curves were also produced from the same data. This is a simpler approach than those described above, and makes no correction for incompletely known taxa.

### *Fossil record completeness*

To assess whether our results may be influenced, and hence explained, by shifts in the quality of the lungfish fossil record, we wish to estimate the completeness of this record through time. A straightforward and powerful way of achieving this is using the simple completeness metric (SCM) of Benton (1987), itself based on earlier work by Paul (1982).

Essentially, this method asks, ‘given that we know a taxon existed at a particular time, how consistently is it sampled?’ This is accomplished by taking the known (observed) range (between its oldest and youngest fossil occurrence, in time bins) of a higher taxon (here, the family) and counting the number of time bins in which a species belonging to that taxon is sampled. For example, if a family spans four time bins and it is sampled in three then it has an SCM of  $3/4 = 75\%$ , or 0.75. This can be considered a Q-mode approach, but the same calculation can be done in R-mode, such that the SCM pertains to a time bin value. So, for example, if 5 families have both first occurrences in or prior to that time bin, and last occurrences in or after that time bin and 4 of them are actually sampled within that bin then the bin has an SCM of  $4/5 = 80\%$ , or 0.8. We make one small modification to the approach of Paul (1982) and Benton (1987): where they include the first and last time bins in the count, we exclude them. This is because they can artificially inflate apparent completeness. For example, a taxon that in reality spans ten bins, if only sampled once, will have an SCM of  $1/1 = 100\%$  or 1, whereas in reality it is  $1/10 = 10\%$  or 0.1. We term our modification SCM\*. To perform this analysis a comprehensive list of lungfish taxa was used, drawing heavily from Schultze (1992) as well as the primary literature. This list has been made available through Dryad [Upload and link will be made available on acceptance].

### *Simulations*

We test the efficacy of our methodology by using a simulated dataset, in which we know the true result (equal rates across the phylogeny) and subsample the idealized tree (simulating an incomplete fossil record) to assess whether our tests still correctly recover the signal of equal rates, or on the contrary, if false positives (the most likely problem with rate analysis) occur. In order to do this we first created the true tree using the following protocol:

1) a random tree of 8000 taxa (near the limits of what is feasible with the desktop computer used) was generated using APE's "rtree" function (Paradis et al. 2004); 2) the randomly assigned branch lengths were converted to millions of years (myr) by making the maximum path-length (root-to-tip length) around 400 myr (the same as the real lungfish data); 3) character changes along a branch were then simulated randomly based on a probability of a change per million years as 0.1 (within the feasible bounds of what we actually observe in the fossil record); and 4) the same changes were assigned to one of 100 different binary characters (again, around the same size as the real lungfish data). This latter point is key to simulating the incompleteness of preservation of individual taxa that we attempt to counter by using Wagner's (1997) patristic dissimilarity.

Once this process was complete the next step was to artificially degrade the data in such a way as to simulate the taphonomic filter of the fossil record. This was done as follows: Firstly, we randomly sampled taxa from the true tree, assuming that a taxon has a probability of 0.01 of actually being sampled. It is difficult to justify a specific value for this figure, but our choice has the advantage of making the sampled tree of similar size to our actual lungfish data. In addition, we feel that we are erring on the side of caution as in reality the sampling rate may not be so bad if comparable to that of other vertebrates such as dinosaurs (Wang & Dodson 2006) or Osteichthyes as a whole (Foote & Sepkoski 1999). Second, we simulated the incompleteness of individual taxa (i.e., the percentage of morphological characters they could be scored for) by randomly choosing a value between 1 and 100% completeness. We allow 100% completeness as, although rare in the fossil record, this is possible, especially if extant taxa are included. On the other hand if a taxon were 100% incomplete it effectively wouldn't be sampled at all. Although these bounds are easily justified, sampling from a uniform distribution probably again errs on the side of caution. Our real data has over 80% of the taxa as at least 40% complete, and in reality chosen taxa and characters will bias the

matrix towards being more complete. Third, the sampled taxa are assigned ages based on randomly sampling species occurrences between the limits of their known true age. We set the number of occurrences to randomly vary between 1 and 8, the best and worst case scenarios from the real lungfish record (tabulated in the previous section). As we are interested in calculating branch durations, the key occurrence here is the earliest sampled date. This part of the simulation is key to understanding how misdating the tree (due to first occurrences of a taxon being long after its first real appearance) can affect rates. At this stage the key input values are available to perform our four rate methods (three heterogeneity tests and temporal rate curve) outlined above.

## **Results**

### *Branch randomization test*

The results of the branch randomizations are shown in Figure 1. Roughly two-thirds of the branches (112 of 170 [66%] for ACCTAN, 109 of 170 [64%] for DELTRAN) show significantly higher rates than expected in at least one of the 1,000 dated trees, indicating a high level of heterogeneity in rates across the phylogeny. However, some branches (67 of 170 [39%] for ACCTAN, 57 of 170 [34%] for DELTRAN) are sensitive to variations in dating, and exhibit a mixture of significant and non-significant results when dates are randomly varied within known limits. General congruence between the two optimizations is moderate; only 78 of 170 (46%) branches give exactly the same results across both optimizations. Of the 92 where exact proportions differ, 9 (5%) give entirely opposite results (all significant versus all non-significant), 41 (24%) differ in absolute proportion but share the most frequent occurrence (i.e., mostly significant or mostly non-significant) and the remaining 42 (25%)

have conflicting frequencies of significant and non-significant results. In summary, although strict congruence across all variations in optimization and dating is low, when a more relaxed view is taken about two-thirds of branches give the same overall picture. By binning these results into Devonian (the traditional rapid phase of lungfish evolution) and post-Devonian (the traditional slow phase) it is clear that higher rates occur more frequently in the Devonian (ACCTTRAN:  $\chi^2 = 8059$ ,  $df = 1$ ,  $p < 10^{-15}$ ; DELTRAN:  $\chi^2 = 3184$ ,  $df = 1$ ,  $p < 10^{-15}$ ). This pattern is consistent across all MPTs assessed (supporting information, table 1).

### *Branch likelihood test*

The results of the branch likelihood test are shown in Figure 2. Under both ACCTTRAN and DELTRAN, the null hypothesis that all branches have equal underlying rates of evolution was conclusively rejected (ACCTTRAN:  $\chi^2 = 1699$ ,  $df = 159$ ,  $p < 10^{-10}$ ; DELTRAN:  $\chi^2 = 1439$ ,  $df = 159$ ,  $p < 10^{-10}$ ). In total 141 of 170 (83%) branches under ACCTTRAN and 147 of 170 (86%) branches under DELTRAN show significant excursions from the one-parameter model (at  $\alpha = .01$  and incorporating the correction for multiple comparisons) on at least one dated tree, again indicating a highly heterogeneous distribution of rates. This time a far greater sensitivity to dating was shown, with differing results between dated trees found on 85 of 170 (50%) branches under ACCTTRAN and 84 of 170 (49%) branches under DELTRAN. Again, congruence between the two optimizations is good – in no cases was the most frequent result in direct opposition (i.e., mostly significantly high on one optimization versus mostly significantly low on the other). Although a lower proportion of branches have identical results across all optimizations and dates (57 of 170, or 34%) a clear signal emerges in their distribution, with most unambiguously high rates found in the Devonian and all unambiguously low rates found in the post-Devonian.

### *Clade likelihood test*

The results of the clade likelihood tests are shown in Figure 3. Most clades show a significant excursion from the one-parameter model (69 of 84 [82%] under ACCTRAN, 77 of 84 [92%] under DELTRAN, at  $\alpha = .01$  and incorporating the correction for multiple comparisons) on at least one dated tree. Of these, a significantly higher rate within the clade (23 of 84 [27%] under ACCTRAN, 33 of 84 [39%] under DELTRAN) is less common than a significantly lower rate (46 of 84 [55%] under ACCTRAN, 44 of 84 [52%] under DELTRAN) when any dated tree is considered. However, 29 of 84 (35%) clades under ACCTRAN and 34 of 84 (40%) under DELTRAN show a sensitivity to dating, with the proportion of significantly high, low and non-significant values varying across dated trees. Similarly, 41 of 84 clades have differing proportions of significantly high, significantly low, and non-significant values between optimisations. Of these 30 (73%) still share the same most frequent occurrence and in none of the other 11 cases were the most frequent values in direct conflict (i.e., the most frequent result was never significantly high for one optimization and significantly low for the other). Consequently, although only 44 of 84 (52%) clades give completely identical proportions across all optimisations and dated trees, overall congruence is still very high. These results thus clearly support the pattern evident in Figure 3 of a persistent slowdown in rates along the backbone of the tree (the path leading to the extant taxa), with clades that show higher rates representing separate, and extinct, excursions.

### *Character changes per lineage million years through time*

Character changes per lineage million years through time are shown in Figure 4. A



clear and pronounced pattern of concentrated character change in the early history of lungfish is notable, and is near identical between the two optimization procedures. Three other minor peaks in rate also occur: in the mid-Carboniferous, just prior to the Permo-Triassic boundary, and in the mid-Cretaceous.

#### *Fossil record completeness*

The results of the SCM\* test are shown in Figure 5. Although they do show a peak value of 1 in the Devonian, when character changes per lineage million years are highest, the same peak is reached in the Carboniferous, Permian, and the Neogene, all times of much lower rates of morphological change. The worst part of the record is in the Late Permian where the SCM\* drops to zero (although there is also a gap at the Permo-Triassic boundary where no data is available). Therefore, based on comparison of the SCM\* and morphological rates curves, it is clear that intervals of high evolutionary rate are not always intervals of either high or low sampling.

To quantitatively assess whether uneven sampling is contributing to, or perhaps driving, our recovered temporal pattern of morphological rates, we correlated the mean SCM\* with the mean number of character changes per lineage million years for each time bin (Figure 4). To avoid problems of autocorrelation we use generalized differencing (McKinney 1990). A very weak positive relationship between SCM\* and the amount of character changes per lineage million years occurs (Spearman  $\rho = 0.08$  and  $0.12$  for ACCTRAN and DELTRAN results, respectively) that is non significant ( $p = 0.64$  and  $0.49$ , respectively). When cross-correlation is applied, to assess whether a stronger lagged relationship exists between the time series, no significantly stronger relationship was found. We therefore conclude that the gross (in)completeness of the lungfish record is not strongly influencing our

results, and that our recovered pattern of morphological rates is generally robust in the face of uneven sampling.

*Simulations: Validity and power of the likelihood-based tests*

Because the branch and clade likelihood tests rely on an asymptotic chi-square approximation (i.e., the test statistic approaches a chi-square distribution as the sample size increases to infinity), we ran simulations to verify the validity of this approximation for a sample size equal to that of our data. Here sample size refers to the number of non-zero-length branches in the tree, equal to 132 in our data. The asymptotic chi-square distribution is exact if the data (the number of changes in each branch) follow a normal distribution. Here we model the number of changes as Poisson rather than normal, with its mean depending on the time duration, completeness, and the rate of change of each branch. However, the Poisson distribution approaches a normal distribution as its mean increases. Thus we expect the asymptotic chi-square approximation to hold better for longer time durations, higher completeness, and higher rates of change. Even when these parameters are not high (and the data as a result are not normal), the chi-square approximation will still hold if the number of branches is sufficiently large (analogous to the Central Limit Theorem). We sought to determine whether the combination of parameter values and sample size was sufficient in our case for the asymptotic chi-square approximation to be valid.

In running the simulations, we considered the tree topology, branch durations ( $t_i$ ), and completeness values ( $c_i$ ) to be fixed, and assigned a true rate of change ( $\lambda_i$ ) uniformly across all branches of the tree. We then simulated a random number of changes for each branch from a  $\text{Poisson}(\lambda_i t_i c_i)$  distribution and carried out the branch and clade likelihood ratio tests. These simulations confirmed that the tests perform as expected (Electronic Supplementary

Figures 7-9). We found that, as expected, the approximation was poor if  $\lambda_i$  was very low (e.g., .01 changes per million years), but the results were reasonable when  $\lambda_i > .08$  changes per million years. Our actual dataset has  $\lambda_i = .148$  changes per million years, well above the value needed for the asymptotic approximation to hold given our tree topology, branch durations, and specimen-level completeness.

We also ran simulations to assess the power of the branch likelihood tests. We found that the test had high power when a branch has a value of  $\lambda_i$  that is substantially higher than the rest of the tree, but low power when  $\lambda_i$  is lower than the rest of the tree (see Electronic Supplementary Material). This is a result of the generally low rate of change in the lungfish tree. Because the overall rate of change is low to begin with ( $\lambda_i = .148$  changes per million years), the expected number of changes on many branches is low, and thus it is difficult to find significant evidence for a lower-than-expected rate of change. (It is for this reason that we used the branch randomization test to detect only branches with significantly elevated rates of change. For trees with higher rates of change, we would not expect a similar problem (and the same problem should not occur for high rates of change, since the number of observed changes in theory has no upper bound).

#### *Simulations: Sensitivity to incompleteness*

We also carried out simulations to assess the sensitivity of our method to the incompleteness of the fossil record. We applied each of the three tests to the equal-rates artificial data that was degraded to represent a poor fossil record. Although two of the three tree-based methods did exhibit instances of Type I errors in the degraded simulations (false positives, i.e., significantly high or low rates when no significant differences should have been found; Electronic Supplementary Figures 2-6), in most cases branches and clades

showed either weakly equivocal (with respect to dating uncertainty) or no excursion from the reality of equal rates across the tree in the undegraded dataset. Using only the results which were unequivocal with respect to dating uncertainty, there were  $33/156 = 21.2\%$  branches with significantly high rates using the randomization test, zero branches with significantly low or significantly high rates using the branch-based likelihood test, and  $8/77 = 10.4\%$  and  $2/77 = 2.6\%$  clades with significantly low and significantly high rates, respectively, using the clade-based likelihood test. The two likelihood-based methods appear to be less likely to return false positives due to the adjustment for multiple comparisons.

Finally, the changes per million years over time data are shown in Electronic Supplementary Figure 5. The simulated data understandably clumps around the preset rate of 0.1 changes per million years, although there is an edge effect due to smaller sample sizes that is exemplified by larger confidence intervals in the earliest and latest parts of the record. The degraded data shows a large missing portion early on (where no taxa were sampled). Otherwise the majority of points oscillate around the preset (true) rate, although there are also several time bins where the confidence intervals do not incorporate this value. In addition, there is one clear outlier at around 90 Ma. The fact that the majority of these excursions are overestimates indicates the underestimation of branch duration (the denominator in the rate equation), most likely due to incomplete sampling.

## **Discussion**

### *Generalities of the method*

Beginning with the influential work of G.G. Simpson (1944), evolutionary biologists have long sought to quantify the tempo of evolutionary change. In particular, paleontologists have offered a deep time perspective on one of the most general but interesting questions in

evolutionary biology: how has organismal morphology changed over time? Many early attempts to explain the pace of morphological evolution were largely narrative-based, but more recently studies have focused on quantifiable morphological features analyzed within a rigorous statistical framework. The majority of these studies examine continuous characters such as body size, but here we use new methods to examine rates of change of discrete characters. Most previous work on discrete character rates has concentrated on changes in rate over time (Wagner 1997; Ruta et al. 2006; Brusatte et al. 2008)—in other words, whether different time intervals have faster or slower rates than others—but our focus is on whether different branches or clades on a phylogeny have heterogeneous rates.

An analogous research trend characterizes the past three decades of biodiversity studies. Initial research focused on quantifying organismal diversity over time (Sepkoski 1978, 1981, 1982; Raup & Sepkoski 1982), but more recently workers have become interested in a separate but related question: are certain clades on the tree of life significantly more diverse than others? More to the point, how can we identify groups that are significantly more diverse than expected by random chance, and what evolutionary phenomena may explain these patterns (Chan & Moore 2002, 2005; McConway & Sims 2004; Moore & Donoghue 2009)? Recently, Chan & Moore (2005) developed a software program, SymmeTREE, that assesses the shape of a phylogeny to test for significant shifts in the rate of diversification (the interplay of speciation and extinction of lineages), and this has been used to identify diversification rate shifts in living (e.g., Jones et al. 2005) and fossil (e.g., Ruta et al. 2007, Lloyd et al. 2008, Tsuji & Muller 2009; Brusatte et al. 2011) taxa. In effect, the clade likelihood test that we present here is a morphological equivalent to SymmeTREE. The underlying question we are testing is the same, except relevant to morphological rather than diversification rates.

The development of an analogous morphological rate test may allow for easier

comparison of rates of diversification and morphological evolution, which are held to be correlated by the most prominent explanations for the theory of punctuated equilibrium (Eldredge & Gould 1972) and may be expected to be correlated during episodes of adaptive radiation (e.g., Schuller 2000; Adams et al. 2009; Gavrilovs & Losos 2009). The first rigorous comparisons between these two rates were just recently undertaken (Adams et al. 2009), but the development of a morphological equivalent to SymmeTREE may allow future workers to compare rates of diversification and morphological change, within a phylogenetic context and based on the same initial dataset (discrete phylogenetic characters and the resulting cladograms).

Additionally, tree-based morphological tests may allow for the identification of heritable rate shifts, or places on the tree where higher or lower rates, later inherited by descendant taxa, were first developed. Identifying such shifts is not foolproof, but the combination of branch and clade tests provides a valuable tool. It is tempting to consider a clade with a higher/lower rate of change than the rest of the tree as one in which there was a shift at its base. However, since the rate of each clade is taken as an average of its constituent branches, it is possible for a small number of extremely high/low branches to drive the whole-clade pattern. With this in mind, if an entire clade is identified as having a significantly high rate and all, or most, of its component branches are significantly high, the hypothesis of a basal, heritable rate shift is tenable. On the contrary, identification of only a few high constituent branches and many low or non-significant branches argues against a heritable shift. In essence, the combination of lineage and clade tests allows for rates to be coarsely optimized onto the tree, just like discrete or continuous morphological characters, and shifts (gains and losses) in rate to be determined. We do not explore this method further for lungfish, as most of the major questions we are trying to address deal with rate heterogeneity over time, but these methods may be useful if researchers are interested in

whether rate shifts coincide with diversification shifts or if rate shifts may be associated with the development of key innovations or the exploitation of new ecological opportunities.

### *Lungfish evolution*

That lungfish have undergone relatively little morphological change subsequent to their first appearance and initial radiation in the Devonian has been noted since Darwin's day and the discovery of the first living lungfish, *Lepidosiren* (Fitzinger 1837) from South America. Perhaps the earliest attempt to trace their lineage was that of Dollo (1895), but calculations of rate would wait until the establishment of the first absolutely dated geologic time scales. This process began in the early part of the 20<sup>th</sup> century with the first radiometric dates (Knopf 1949). Simpson (1944) is widely regarded to have pioneered the study of evolutionary rates, placing time as palaeontology's key contribution to the wider field of evolutionary biology. However, although landmark studies, most of Simpson's methods were based on change in a single continuous character, such as body size or tooth shape.

The natural starting place for the present contribution is the study of Westoll (1949). This was one of the earliest attempts to gauge the rate of evolutionary change across the whole organism, and like our study, focused on lungfish. Using 26 characters, Westoll gave 15 species scores ranging from 100 (hypothetical ancestor) to 0 (the living lepidosirenids). When plotted against time, it was clear that the largest change in score occurred in a relatively short period of time early in lungfish history and that subsequent change was indeed slow. Simpson (1953) assimilated Westoll's results, reversing the numbering so that it reflected an increase in score through time, emphasising the acquisition of novel characters instead of the loss of primitive ones, as well as calculating absolute rates. Our figure 4 is thus an updated version of Simpson's (1953) figure 4.

Following Simpsons's (1944, 1953) and Westoll's (1949) results, lungfish became a classic example of "living fossils" and were often used to illustrate general principles of evolution in textbooks and popular accounts. Surprisingly, large-scale patterns of lungfish evolution were not revisited again until Wagner (1980) attempted to fit different models of character change to Westoll's original results. Subsequently, lungfish evolution has been revisited twice with more recent datasets (Martin 1987; Schultze 2004). Martin (1987) was the first author to note that ancestor-descendant pathways leading to different taxa give slightly differing results (in his figure 10 Triassic ceratodontids had a faster, and Triassic ptychoceratodontids a slower, rate than the extant lepidosirenids). However, Westoll's (1949) most important conclusion, that character evolution was rapid in the Devonian but followed by slower rates of change, was still upheld. Schultze (2004) used a version of the present data set to calculate rates over fairly coarse bins (equal or greater in length than a geological Period), grouping all dipnoans together (rather than showing differences between branches of the tree) to give an average "rate of morphological evolution" through time. Again these results upheld Westoll's original conclusions.

Here, figure 4 represents the best comparison to these earlier efforts. Globally the picture is congruent with these earlier studies; however, hitherto unrecognized minor increases in rates are observable, most prominently at around 255 Ma, roughly coincident with the Permo-Triassic boundary and is present across all MPTs assessed. We suspect that this pattern is only absent from earlier studies due to their coarser time bins. Forey (1988) also found the highest rate of coelacanth evolution, and greatest number of taxa, to occur nearby, in the Triassic, indicating this period may have been as globally important for the evolution of non-tetrapod sarcopterygians as it was for tetrapods (Sahney & Benton 2008). In other words, the Permo-Triassic may have been a critical stage in the evolution of lungfish, a pattern that is just beginning to emerge due to larger phylogenetic datasets, better absolute



dating control, and more sophisticated rate methods.

The present study, based on an expanded character dataset and three novel methods of assessing rate heterogeneity across a phylogenetic tree, also adds several other new observations to the tempo of lungfish evolution. The first of these (figure 1) shows a decrease in significantly high rates over time. It also suggests that a null hypothesis of equal rates across the tree is a very poor model for what is observed in the fossil record, with most branches representing a significant excursion to high rates across most optimizations and dating randomizations.

In the likelihood approach, although most branches are again significantly different than expectations, the excursions are more symmetrically distributed (between higher and lower rates). Furthermore, the ability to distinguish low as well as high rates is a strength of the likelihood approach. From these results, it is clear that the early backbone of the tree is dominated by high rates, but in addition that this pattern actually continues into the post-Devonian. Indeed, the first branch on the line to the extant taxa that shows any significantly lower rates subtends the *Gnathorhiza-Lepidosiren* clade, dated as Late Carboniferous (Pennsylvanian) in age. However, even here the majority of branches are still non-significant. In fact only one branch on the direct pathway to the three extant genera is significantly low across all optimisations and date randomizations. This is the branch subtending the crown group (containing the three extant genera, plus *Mioceratodus*), dated as Early Cretaceous in age. Such results offer a more complex picture than offered by previous workers, although Schultze (2004) did show a rise in rates in the Cretaceous and Martin's (1987) pattern is not as smooth as Westoll's (1949). However, neither worker suggested that the major tapering off of evolutionary rates on the line to the extant taxa was actually postponed until the late Mesozoic, which our data seem to indicate.

The reason for the seeming incongruence between this result and those shown in

figure 4 is twofold. First, the approach used in figure 4 (and by Schultze 2004) groups all changes by time, regardless of the number of branches involved, and branches are surely more numerous in the Devonian than post-Devonian. Indeed, the relative size of the Devonian and post-Devonian portions of the tree in figures 1-3 confirms this.

Consequently, misinterpretation of figure 4 could lead to the expectation that per-branch rates ought to be uniformly lower in the post-Devonian. However, the results of Martin (1987), where more post-Devonian than Devonian branches are sampled, do seem to show low rates (shallow slopes) on three separate lineages in the post-Devonian. Closer inspection, though, does show one branch (from *Tellerodus formosus* to *Tellerodus sturii*, at the termination of the “Triassic ceratodontids”, his figure 10) to be significantly steeper and the lepidosirenid-line is not broken-up (i.e., taxonomically sampled) to the same degree as it is here. A second reason for the difference is that the phylogenetic methods used here include an important additional correction for sampling bias (incompleteness of individual taxa: patristic dissimilarity, Wagner 1997), whereas the approaches of figure 4 and previous authors either require or assume complete or near complete preservation of each taxon. Taken together, the results of the branch-based likelihood approach argue for the phylogenetic elucidation of patterns that a “binning-by-time” method obscures.

Our final phylogenetic approach compares entire clades. Here a very clear pattern is apparent (figure 3), seemingly indicating that the low rates present in the crown group permeate throughout all clade comparisons, such that any clade including the crown will be shown to be evolving slower than the rest of the tree. The corollary to this is that the three large (five or more taxa) Devonian clades that diverge from the extant-line show relatively high rates. However, the three post-Devonian clades (*Tranodis-Ctenodus*, *Parasagenodus-Ceratodus latissimus* and *Microceratodus-Arganodus*) divergent from the extant line do not. Consequently, the clade-based likelihood approach seems to revert to a similar, more

simplistic, interpretation as that of figure 4 and earlier studies, namely that Devonian rates are high and post-Devonian rates are slow. However, our methods take into account sampling biases, in terms of number of branches, time duration of branches, and preservation of taxa putting the coarser-level interpretation of previous studies on a sounder footing.

In summary these results suggest that lungfish morphological evolution has indeed slowed down since the Devonian, but notable exceptions occur when the data are viewed at a higher resolution, either in terms of shorter time bins or at the level of individual branches. Most significantly, living lungfish may not have seen a significant retardation of the rate of morphological evolution in their ancestral line until the Cretaceous, rather than the Palaeozoic as previously thought. In other words, lungfish may have dwindled to “living fossil” status much more recently, and have not been stagnant for over 300 million years as the textbook story holds.

## **Acknowledgements**

The authors thank Hans-Peter Schultze for sharing the matrix and character list from Schultze and Marshall (1993), Mike Benton, Phil Donoghue, Gene Hunt, Marcello Ruta, and Lee Hsiang Liow as well as an anonymous reviewer for helpful comments on earlier drafts of this manuscript and Marcello Ruta, Manabu Sakamoto, and Phil Everson for discussions. For the instruction in R received as part of the Paleobiology Database Intensive Summer Course in Analytical Paleobiology (<http://paleodb.org>) GTL would additionally like to thank John Alroy, Michael Foote, Tom Olszewski, David Polly and Pete Wagner. GTL was supported by NERC studentship NER/S/A/2004/12222. SCW is grateful for partial support from the Swarthmore College Research Fund. SLB is supported by an NSF Graduate Research Fellowship, NSF Doctoral Dissertation Improvement Grant (NSF DEB 1110357), and his research has also been supported by the Marshall Scholarship for study in the UK.

## References

- Adams DC, Berns CM, Kozak KH, Wiens JJ. Are rates of species diversification correlated with rates of morphological evolution? *P Roy Soc B*. 2009;276-2729-2738.
- Ahlberg PE, Smith MM. Developmental plasticity and disparity in early dipnoan (lungfish) dentitions. *Evol Dev*. 2006;8:331–349.
- Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Nat. Acad. Sci. (USA)* 2009;106:13410-13414.
- Alroy J, Aberhan M, Bottjer DJ, Foote M, Fursich F T, Harries PJ, Hendy AJW, Holland SM, Ivany LC, Kiessling W, Kosnik MA, Marshall CR, McGowan AJ, Miller AI, Olszewski TD, Patzkowsky ME, Peters SE, Villier L, Wagner PJ, Bonuso N, Borkow PS, Brenneis B, Clapham ME, Fall LM, Ferguson CA, Hanson VL, Krug AZ, Layou KM, Leckey EH, Nurnberg S, Powers CM, Sessa JA, Simpson C, Tomasovych A, Visaggi CC. Phanerozoic trends in the global diversity of marine invertebrates. *Science* 2008;321:97–100.
- Apesteguia S, Agnolin FL, Claeson K. Review of Cretaceous dipnoan from Argentina (Sarcopterygii, Dipnoi) with descriptions of new species. *Rev Mus Argentino Cienc Nat*, ns. 2007;9:27-40.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B*. 1995;57:289–300.
- Benton MJ. Mass extinctions among families of non marine tetrapods: the data, *Mém. Soc géol France* n.s. 1987;150:21-32.

- Bininda-Emonds ORP. perlRat.pl v1.0.9a. Program distributed by author. Institut für Spezielle Zoologie und Evolutionsbiologie mit Phyletischem Museum, Friedrich-Schiller-Universität Jena; 2004.
- Bromham L, Woolfit M, Lee MSY, Rambaut A. Testing the relationship between morphological and molecular rates along phylogenies. *Evolution* 2002;56:1921-1930.
- Brusatte SL. Calculating the tempo of morphological evolution: rates of discrete character change in a phylogenetic context. In: Elewa, AMT, editor. *Computational Paleontology*. Heidelberg : Springer-Verlag; 2011. p. 53-74.
- Brusatte SL, Benton MJ, Lloyd GT, Ruta M, Wang SC. Macroevolutionary patterns in the evolutionary radiation of archosaurs (Tetrapoda: Diapsida). *Earth Env. Sci. Trans. R. Soc. Edin.* 2011;101:367-382.
- Brusatte SL, Benton MJ, Ruta M, Lloyd GT. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science* 2008;321:1485-1488.
- Cavin L, Suteethorn V, Buffetaut E, Tong H. A new Thai Mesozoic lungfish (Sarcopterygii, Dipnoi) with an insight into post-Palaeozoic dipnoan evolution. *Zool J Linn Soc.* 2007;149:141–177.
- Chan KMA, Moore BR. Whole-tree methods for detecting differential diversification rates. *Syst Biol* 2002;51: 855-865.
- Chan KMA, Moore BR. SymmeTREE: Whole-tree analysis of differential diversification rates. *Bioinformatics* 2005;21:1709-1710.
- Chaloner, WG, Sheerin A. Devonian macrofloras. *Spec Pap Palaeo.* 1979;23:145–161.
- Cloutier R. Patterns, trends, and rates of evolution within the Actinistia. *Env Biol Fish.* 1991;32:23–58.
- Collar DC, Near TJ, Wainwright PC. Comparative analysis of morphological diversity: does

disparity accumulate at the same rate in two lineages of centrarchid fishes? *Evolution* 2005;59:1783-1794.

Cooper N, Purvis A. What factors shape rates of phenotypic evolution? A comparative study of cranial morphology of four mammalian clades. *J Evol Biol* 2009;22: 1024-1035.

Darwin C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: J. Murray; 1859.

Dawkins R. The Ancestor's Tale: A Pilgrimage to the Dawn of Life. London: Weidenfeld Nicolson Illustrated; 2004.

Derstler KL. Estimating the rate of morphological change in fossil groups. *Proc Third N Amer Paleo Conv.* 1982;1:131–136.

Dollo L. Sur la phylogenie des dipneustes. *Bull Soc belge Geol Pal Hydr.* 1895;979-128.

Eldredge N, Gould SJ. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM, editor. *Models in Paleobiology*. San Francisco: Freeman, Cooper and Co.; 1972. p. 92–115.

Fitzinger LJ. Über *Lepidosiren*. *Archiv Naturgesch.* 1837;2:232.

Foote M, Sepkoski JJ. Absolute measures of the completeness of the fossil record. *Nature.* 1999;398:415-417.

Forey PL. Golden Jubilee for the coelacanth *Latimeria chalumnae*. *Nature* 1988;336:727–732.

Friedman M. The interrelationships of Devonian lungfishes (Sarcopterygii: Dipnoi) as inferred from neurocranial evidence and new data from the genus *Soederberghia* Lehman, 1959. *Zool J Linn Soc.* 2007;151:115–171.

Garland TG. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am Nat* 1992;140:509-519.

- Gatesy J, Matthee C, deSalle R, Hayashi C. Resolution of a supertree-supermatrix paradox. *Syst Biol.* 2002;51:652-664.
- Gatesy J, Baker RH, Hayashi, C. Inconsistencies in arguments for the supertree approach: supermatrices versus supertrees of Crocodylia. *Syst Biol.* 2004;53:342–355.
- Gavrilets S, Losos JB. Adaptive radiation: contrasting theory with data. *Science* 2009;323:732-737.
- Gradstein FM, Ogg JG, Smith AG. A Geologic Time Scale 2004. Cambridge: Cambridge University Press; 2004.
- Harmon LJ, Melville J, Larson A, Losos JB. The role of geography and ecological opportunity in the diversification of day geckos (*Phelsuma*). *Syst Biol* 2008;57:562-573.
- Hill RV. Integration of morphological data sets for phylogenetic analysis of Amniota: the importance of integumentary characters and increased taxonomic sampling. *Syst Biol.* 2005;54:530-547.
- Hutcheon JM, Garland T. Are megabats big? *J Mammal. Evol.* 2004;11:257-277.
- Jones S. Almost Like a Whale. London: Doubleday; 1999.
- Jones KE, Bininda-Emonds ORP, Gittleman JL. Bats, clocks, and rocks: diversification patterns in Chiroptera. *Evolution* 2005;59:2243–2255.
- Knopf A. Time in Earth history. In: Jepsen GL, Simpson GG, Mayr E, editors. *Genetics, Paleontology and Evolution*. Princeton: Princeton University Press; 1949. p. 1-9.
- Kumar S, Hedges SB. A molecular timescale for vertebrate evolution. *Nature* 1998;392:917-920.
- Krupina NI, Reisz RR. Phylogeny of Devonian dipnoans, with special emphasis on forms from the Russian platform. *Paleontological Journal* 2000;34:S147–S153.
- Levinton JS. *Genetics, Paleontology, and Macroevolution* (2nd Ed.). Cambridge: Cambridge University Press; 2001.

- Lloyd GT, Davis KE, Pisani D, Tarver JE, Ruta M, Sakamoto M, Hone DWE, Jennings R, Benton MJ. Dinosaurs and the Cretaceous Terrestrial Revolution. *Proc Roy Soc B*. 2008;275:2483-2490.
- Marshall CR. Lungfish: phylogeny and parsimony. *J Morph Supp*. 1986;1:151–162.
- Martin M. Dipnoan evolution revisited. *Mém Soc géol France NS*. 197;150:13–20.
- McConway KJ, Sims HJ. A likelihood-based method for testing for non-stochastic variation of diversification rates in phylogenies. *Evolution* 2004;58:12-23.
- McKinney ML. Classifying and analyzing evolutionary trends. In: McNamara KJ, editor. *Evolutionary Trends*. London: Belhaven; 1990. P. 28-58.
- Mindell DP, Thacker CE. Rates of molecular evolution: phylogenetic issues and applications. *Ann Rev Ecol Syst* 1996;27:279-303.
- Moore BR, Donoghue MJ. A Bayesian approach for evaluating the impact of historical events on rates of diversification. *P Natl Acad Sci (USA)* 2009;106: 4307-4312.
- Nixon KC. The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics* 1999;15:407–414.
- Oakley TH. Maximum likelihood models of trait evolution. *Comments on Theoretical Biology* 2003;8:1-17.
- O’Meara BC, Ané C, Sanderson MJ, Wainwright PC. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 2006;60:922-933.
- Omland K. Correlated rates of molecular and morphological evolution. *Evolution* 1997;51:1381-1393.
- Pagel M. Inferring evolutionary process from phylogenies. *Zoologica Scripta* 1998;26:331-348.
- Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20:289–290.



- Paul, CRC. The adequacy of the fossil record. In: Joysey KA, Friday AE, editors. Problems of Phylogenetic Reconstruction. London: Academic Press; 1982. P. 75-117.
- Pinto G, Mahler DL, Harmon LJ, Losos JB. Testing the island effect in adaptive radiation: rates and patterns of morphological diversification in Caribbean and mainland *Anolis* lizards. *P Roy Soc B* 2008;275:2749-2757.
- Pol D, Norell, MA. Uncertainty in the age of fossils and the stratigraphic fit to phylogenies. *Syst Biol.* 2006;55:512-521.
- Qiao T, Zhu M. A new tooth-plated lungfish from the Middle Devonian of Yunnan, China, and its phylogenetic relationships. *Acta Zool.* 2009;90(supp. 1):236-252.
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>; 2010.
- Raup DM, Sepkoski JJ. Mass extinctions in the marine fossil record. *Science* 1982;215:1501–1503.
- Raup DM, Stanley SM. Principles of Paleontology (2nd Ed.). San Francisco: W. H. Freeman and Company; 1978.
- Ree RH. Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution* 2005;59:257-265.
- Revell LJ, Harmon LJ, Collar DC. Phylogenetic signal, evolutionary process, and rate. *Syst Biol* 2008;57:591-601.
- Ricklefs RE. Estimating diversification rates from phylogenetic information. *Trends Ecol Evol* 2007;22:601-610.
- Ridley M. Evolution (2nd Ed.). Oxford: Blackwell Science; 1996.

- Ruta M, Wagner PJ, Coates MI. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proc Roy Soc B*. 2006;273:2107–2111.
- Ruta M, Pisani D, Lloyd GT, Benton MJ. A supertree of Temnospondyli: cladogenetic patterns in the most species-rich group of early tetrapods. *Proc Roy Soc B*. 2007;274:3087-3095.
- Sahney S, Benton MJ. Recovery from the most profound mass extinction of all time. *Proc Roy Soc B*. 2008;275:759-765.
- Schultze H-P. *Fossilium Catalogus I: Animalia pars 131 Dipnoi*. Amsterdam: Kugler; 1992.
- Schultze H-P. *Melanognathus*, a primitive dipnoan from the Lower Devonian of the Canadian Arctic and the interrelationships of Devonian dipnoans. *J Vert Paleo*. 2001;21:781–794.
- Schultze H-P. Mesozoic sarcopterygians. In: Arratia G, Tintori, A, editor. *Mesozoic Fishes 3 - Systematics, Paleoenvironments and Biodiversity*. Munchen: Verlag Dr. Freidrich Pfeil; 2004. p. 463–492.
- Schultze H-P, Chorn J. The Permo-Carboniferous genus *Sagenodus* and the beginning of modern lungfish. *Cont Zool*. 1997;67:9–70.
- Schultze H-P, Marshall CR. Contrasting the use of functional complexes and isolated characters in lungfish evolution. *Mem Assoc Austral Palaeo*. 1993;15:211-224.
- Sepkoski JJ. A kinetic model of Phanerozoic taxonomic diversity, I: Analysis of Marine Orders. *Paleobiology* 1978;4:223–251.
- Sepkoski JJ. A factor analytic description of the Phanerozoic marine fossil record. *Paleobiology* 1981;7:36–53.
- Sepkoski JJ. A compendium of fossil marine families. *Milw Pub Mus Cont Biol Geol*. 1982;51:1-156.
- Sidlauskas BL. Testing for unequal rates of morphological diversification in the absence of a

- detailed phylogeny: a case study from characiform fishes. *Evolution* 2007;61:299-316.
- Simpson GG. *Tempo and Mode in Evolution*. New York: Columbia University Press; 1944.
- Simpson GG. *The Major Features of Evolution*. New York: Simon and Schuster; 1953.
- Sims HJ, McConway KJ. Non-stochastic variation of species-level diversification rates within angiosperms. *Evolution* 2003;57:460-479.
- Stanley SM. *Macroevolution: Pattern and Process*. San Francisco: W. H. Freeman and Company; 1979.
- Swofford DL. *PAUP\*: Phylogenetic Analysis Using Parsimony (\*and other methods)*. Sunderland: Sinauer Associates; 2003.
- Swofford DL, Maddison WP. Reconstructing ancestral character states under Wagner parsimony. *Math Biosci.* 1987;87:199–229.
- Thomas GH, Freckleton RP, Székely. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proc Roy Soc B* 2006;273:1619-1624.
- Tsuji LA, Müller J. Assembling the history of the Parareptilia: phylogeny, diversification, and a new definition of the clade. *Fossil Record* 2009;12:71-81.
- Wagner GP. Empirical information about the mechanism of typogenetic evolution. *Naturwissenschaften* 1980;67:258-259.
- Wagner PJ. Patterns of morphologic diversification among the Rostroconchia. *Paleobiology* 1997;23:115–150.
- Wang SC, Dodson P. Estimating the diversity of dinosaurs. *Proc Natl Acad Sci USA*. 2006;103:13601-13605.
- Westoll TS. On the evolution of the Dipnoi. In: Jepsen GL, Simpson GG, Mayr E, editors. *Genetics, Paleontology and Evolution*. Princeton: Princeton University Press; 1949. p. 121–184.

Wilkinson M. Common cladistic information and its consensus representation: reduced Adams and reduced cladistic consensus trees and profiles. *Syst Biol.* 1994;43:343–368.

Young GC, Schultze H-P. New osteichthyans (bony fishes) from the Devonian of Central Australia. *Mitteil Mus Natur Berlin, Geowiss R.* 2005;8:13–35.

**Figure 1.** Branch randomization results. Pie charts indicate the proportional distribution of results across 1,000 dated trees using the sensitivity test described in the text and are split in two, with ACCTRAN results on top and DELTRAN results below in all cases. Green indicates a significantly higher rate than expected ( $p < 0.05$ ; compared to a randomised distribution assuming equal rates) and white a non-significant result. For ease of viewing the figured tree is ladderized and all branch lengths set to one. The three extant genera (*Lepidosiren*, *Neoceratodus* and *Protopterus*) are found at the lowest right. Note the greater number of non-significant results (white) in the post-Devonian (grey box).

**Figure 2.** Branch likelihood test results. Key as in figure 1, but here red indicates a significantly lower rate than expected under the one-parameter model. Note the concentration of red (low rates) and white (non-significant branches) in the post-Devonian (grey box).

**Figure 3.** Clade likelihood test results. Key as in figure 1, except here red indicates a significantly lower rate than expected under the one-parameter model. Values at nodes indicate the rate for the clade subtended by that node compared to the rest of the tree (see text). Note the absence of green (high rates) in the post-Devonian (grey box).

**Figure 4.** Character changes per lineage million years results under a) ACCTRAN and b) DELTRAN. Geologic period abbreviations are as follows: Sil. = Silurian, Devon. = Devonian

and Ng. = Neogene. Results are the composite of 1,000 runs using the sensitivity test described in the text. Points indicate the number of changes per lineage million years that occurred in each bin with error bars showing the 95% confidence intervals. The line shows the same results, but this time as accumulated changes with the shaded area indicating the 95% confidence intervals. This latter approach is the closest way of replicating Westoll's (1949) original lungfish analysis with a cladistic data set. Note the peak in rates at around 400 Ma as well as "kinks" at around 325 (mid-Carboniferous), 250 Ma (the Permo-Triassic boundary), and 100 Ma (mid-Cretaceous).

**Figure 5** Single Completeness Metric\* values for the lungfish fossil record. Solid line indicates mean, and grey the 95% confidence intervals of 1,000 randomisations that account for uncertainty in dating. Abbreviations for geologic periods (along x-axis) are as in figure 4.